การจำแนกมาตรวัดน้ำที่ผิดปกติ โดยใช้เทคนิคการเรียนรู้ของเครื่องร่วมกับข้อมูลเชิงพื้นที่

Classification of Anomalous Water Meters Using Machine Learning and Spatial Data

ภานุวัฒน์ โมหนองเดิ่น[1,*] ชัยยุทธ เจริญผล[2] และ ธุวชิต แฉล้มเขตต์[3]

[1,2] ภาควิชาวิศวกรรมสำรวจ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จ.กรุงเทพมหานคร
[3] การประปาส่วนภูมิภาค จ.กรุงเทพมหานคร
**Corresponding author; E-mail address: Chaiyut@chula.ac.th*

บทคัดย่อ

การบริหารจัดการน้ำสูญเสีย เป็นภารกิจสำคัญของการให้บริการน้ำประปา โดยน้ำสูญเสียเกิดจากการรั่วไหลทางกายภาพของระบบท่อและน้ำสูญเสียเชิงพาณิชย์จากมาตรวัดน้ำที่ชำรุด ซึ่งส่งผลให้การประปาส่วนภูมิภาค (กปภ.) สูญเสียรายได้และกระทบต่อความเชื่อมั่นของผู้ใช้น้ำ การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองในการจำแนกมาตรวัดน้ำที่ผิดปกติโดยใช้การเรียนรู้ของเครื่อง (Machine Learning: ML) ผ่านแบบจำลอง XGBoost, LSTM และ Stacking Ensemble งานวิจัยใช้ข้อมูลอายุของมาตรวัดน้ำ ประเภทผู้ใช้น้ำ ยี่ห้อ ประวัติการใช้น้ำย้อนหลัง 49 เดือน และข้อมูลเชิงพื้นที่ เช่น จำนวนจุดซ่อมท่อ ระยะห่างจากจุดซ่อมท่อในพื้นที่ กปภ. สาขารังสิต จังหวัดปทุมธานี ผลการศึกษาพบว่าแบบจำลอง Stacking Ensemble ที่รวมข้อมูลเชิงพื้นที่ให้ความแม่นยำสูงสุดที่ 90.03% งานวิจัยนี้ช่วยให้ กปภ. คัดกรองมาตรวัดน้ำที่ผิดปกติได้แม่นยำ ลดน้ำสูญเสีย เพิ่มความเป็นธรรมในการคิดค่าน้ำ และสามารถขยายผลสู่พื้นที่อื่นในอนาคต

คำสำคัญ: น้ำสูญเสียเชิงพาณิชย์, มาตรวัดน้ำ, การเรียนรู้ของเครื่อง, ข้อมูลเชิงพื้นที่, การประปาส่วนภูมิภาค

Abstract

Water loss management is a crucial task in water supply services. Water loss occurs due to physical leakage in the pipeline system and commercial losses from malfunctioning water meters, leading to revenue loss and reduced consumer confidence for the Provincial Waterworks Authority (PWA). This study aims to develop a machine learning (ML) model to classify water meters prone to malfunction using XGBoost, LSTM, and a Stacking Ensemble approach. The research utilizes data on meter age, customer type, brand, 49-month historical consumption records, and spatial data such as the number of pipe repair points and the distance to repair sites in the PWA Rangsit branch, Pathum Thani Province. The results show that the Stacking Ensemble model incorporating spatial data achieves the highest accuracy 90.03%. This study enables PWA to accurately detect malfunctioning water meters, reduce water loss, enhance fairness in water billing, and expand model applications to other areas in the future.

Keywords: Commercial water loss, Water meters, Machine learning, Spatial data, Provincial Waterworks Authority (PWA)

1. บทนำ

น้ำเป็นสิ่งสำคัญต่อการต่อการดำรงชีวิตของมนุษย์ทั้งในด้านการอุปโภคและบริโภค ปัญหาการขาดแคลนน้ำและการใช้น้ำอย่างไม่คุ้มค่าเป็นปัญหาที่เกิดขึ้นในหลายภูมิภาคทั่วโลก การบริหารจัดการน้ำอย่างมีประสิทธิภาพจึงเป็นสิ่งจำเป็นในการป้องกันการสูญเสียทรัพยากรที่มีค่า เช่น การรั่วไหลของน้ำจากท่อและความถูกต้องแม่นยำของมาตรวัดน้ำถือเป็นสิ่งสำคัญสำหรับการจัดการน้ำที่มีประสิทธิภาพและการเรียกเก็บเงินที่เป็นธรรม “น้ำสูญเสีย” (Non-Revenue Water: NRW) เป็นหนึ่งปัจจัยที่สะท้อนผลสำเร็จโดยตรงของการบริหารกิจการประปา และเป็นอุปสรรคต่อความยั่งยืนของการจัดการทรัพยากรน้ำในระยะยาว น้ำสูญเสียเป็นน้ำที่จ่ายเข้าระบบแล้วไม่ก่อให้เกิดรายได้ ประกอบด้วย 1) น้ำสูญเสียเชิงกายภาพ (Physical Losses) เกิดจากการรั่วไหลของน้ำในระบบท่อส่งน้ำ 2) น้ำสูญเสียเชิงพาณิชย์ (Commercial Losses) เกิดจากการวัดปริมาณน้ำที่ผิดพลาด การลักใช้น้ำ หรือการคิดค่าน้ำไม่ถูกต้อง 3) น้ำจ่ายเพื่อบริการสาธารณะคือน้ำที่ไม่มีการเรียกเก็บเงิน เช่น การบรรเทาสาธารณภัย การระบายตะกอนในระบบท่อหรือล้างถังน้ำใสซึ่งเป็นการบำรุงรักษาระบบประปา [1]

การประปาส่วนภูมิภาค (กปภ.) เป็นหน่วยงานหลักในการให้บริการสาธารณะด้านการผลิตและบริการน้ำประปาให้แก่ประชาชน ใน 74 จังหวัดของประเทศไทย ยกเว้นกรุงเทพมหานคร สมุทรปราการ และนนทบุรี ซึ่ง

ให้บริการโดยการประปานครหลวง (กปน.) ปัจจุบัน ณ กันยายน 2567 กปภ. มีจำนวนผู้ใช้น้ำกว่า 5.452 ล้านราย มีอัตราเฉลี่ยการใช้น้ำตามประเภทผู้ใช้น้ำได้ดังนี้ บ้านพักอาศัย 533 ลิตร/วัน ธุรกิจขนาดกลาง 1,437 ลิตร/วัน และ สถานที่ราชการหรืออุตสาหกรรมขนาดใหญ่ 3,918 ลิตร/วัน [2]

น้ำสูญเสียเป็นปัญหาสำคัญที่ส่งผลกระทบทางเศรษฐกิจอย่างมหาศาล โดยมีมูลค่าความเสียหายรวมหลายพันล้านบาทต่อปี ด้วยเหตุนี้ การประปาส่วนภูมิภาค (กปภ.) จึงให้ความสำคัญกับการบริหารจัดการน้ำสูญเสีย ซึ่งถือเป็นหนึ่งในปัญหาหลักขององค์กร ปัจจุบันปริมาณน้ำสูญเสียในระบบจำหน่ายยังคงอยู่ในระดับสูง โดยเฉพาะในเขตชุมชนเมือง ซึ่งเป็นพื้นที่ให้บริการของสาขาขนาดใหญ่ที่มีจำนวนผู้ใช้น้ำเป็นจำนวนมาก แม้ว่าจะมีมาตรการลดน้ำสูญเสียเชิงกายภาพ ซึ่งเป็นปัจจัยหลักของปัญหา แต่ปริมาณน้ำสูญเสียก็ยังคงอยู่ในระดับสูง

ดังนั้น กปภ. จึงจำเป็นต้องดำเนินการลดน้ำสูญเสียเชิงพาณิชย์ควบคู่ไปด้วย แม้ว่าสาเหตุนี้จะเป็นปัจจัยรอง แต่การบริหารจัดการที่มีประสิทธิภาพจะช่วยลดปริมาณน้ำสูญเสียได้เพิ่มเติม ทั้งนี้การดำเนินการดังกล่าวจะดำเนินไปพร้อมกับการบำรุงรักษาระบบจ่ายน้ำให้เป็นไปตามมาตรฐานที่กำหนด เพื่อเพิ่มประสิทธิภาพในการให้บริการน้ำประปาและลดผลกระทบจากน้ำสูญเสียในระยะยาว [3]

การศึกษาวิจัยพบว่าการลดน้ำสูญเสียเชิงพาณิชย์ทำได้ง่ายและรวดเร็วกว่าการลดน้ำสูญเสียเชิงกายภาพซึ่งเป็นปัญหาหลัก [4] อีกทั้งยังช่วยเพิ่มรายได้ให้กับ กปภ.โดยในปี 2565 กปภ. เริ่มกำหนดแผนปฏิบัติงานเพื่อลดน้ำสูญเสียเชิงพาณิชย์ (ปรับปรุงล่าสุด 2567) โดยเน้นการค้นหามาตรวัดน้ำที่มีแนวโน้มผิดปกติ เช่น มาตรวัดที่อ่านค่าไม่เที่ยงตรง ค่าคลาดเคลื่อนเกินมาตรฐาน หรืออาจเกิดจากการฉ้อโกง การอ่านค่าผิดพลาด และการรั่วไหลภายในอาคาร [5] บางกรณีแม้มาตรวัดน้ำจะดูปกติ แต่ยังอาจมีความบกพร่องที่ตรวจจับได้ยาก การวัดปริมาณน้ำที่ผิดพลาดส่งผลให้ กปภ. สูญเสียรายได้ หรือสร้างความไม่เป็นธรรมต่อผู้ใช้น้ำ ดังนั้นหากสามารถคัดแยกและเปลี่ยนมาตรวัดน้ำที่ผิดปกติได้ทันเวลา รายได้ที่หายไปจะกลับคืนสู่ระบบ และช่วยเสริมความเชื่อมั่นของผู้ใช้น้ำ โดยในระยะแรก กปภ. ใช้ระบบติดตามมาตรวัดน้ำ (MMS) ควบคู่กับข้อมูล GIS เพื่อตรวจหามาตรวัดน้ำที่ผิดปกติ โดยวิเคราะห์ข้อมูลย้อนหลัง 4 ปี เช่น ค่าเฉลี่ยการใช้น้ำ การเปลี่ยนแปลงรายปี และค่าสัมประสิทธิ์ความแปรผัน [6] อย่างไรก็ตาม พบว่าข้อมูลที่คัดกรองมีข้อผิดพลาดสูง เนื่องจากพฤติกรรมการใช้น้ำของแต่ละรายแตกต่างกัน และอาจมีปัจจัยแฝงที่ระบบยังไม่ครอบคลุม จึงต้องปรับปรุงเงื่อนไขและกระบวนการวิเคราะห์ให้แม่นยำและครอบคลุมยิ่งขึ้น

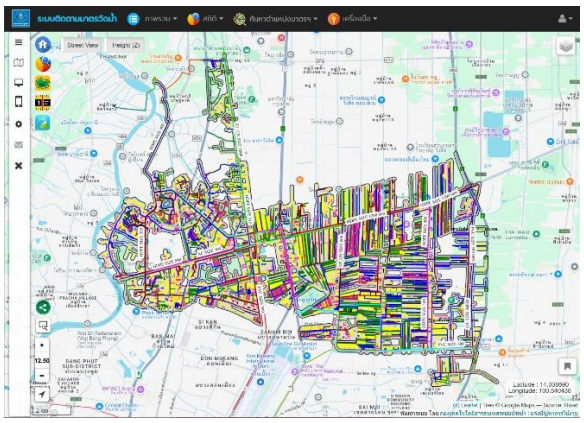
เพื่อลดข้อผิดพลาดในการจำแนกมาตรวัดน้ำผิดปกติ ผู้วิจัยจึงศึกษาวรรณกรรมที่เกี่ยวข้องและพบว่า Machine Learning (ML) เป็นเทคนิคที่ได้รับการประยุกต์ใช้ในการตรวจจับมาตรวัดน้ำที่มีแนวโน้มผิดปกติ เช่น การใช้ Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM) สำหรับตรวจจับการรั่วไหลภายในครัวเรือน [7] การใช้ z-score (ZS), local outlier factor (LOF) one-class support vector machine (OCSVM) และ isolation forest (iForest) สามารถตรวจจับการใช้น้ำที่ผิดปกติจากประวัติการใช้น้ำในอดีตได้ ซึ่งการใช้น้ำที่ไม่ปกติอาจเกิดจากมาตรวัดน้ำที่เสีย การดัดแปลงมาตรวัดน้ำอย่างผิดกฎหมาย หรือการรั่วไหลของน้ำในท่อน้ำภายในอาคารบ้านเรือนของผู้บริโภค [5] และ พบว่าการใช้หลายเทคนิค ML ร่วมกันช่วยเพิ่มความแม่นยำในการตรวจจับความผิดปกติ [8]

อย่างไรก็ตาม งานวิจัยที่เกี่ยวข้องยังขาดการศึกษาเรื่องการตรวจจับมาตรวัดน้ำชำรุดที่แม่นยำ ผู้วิจัยจึงนำเทคนิค ML มาผสานกับข้อมูลเชิงพื้นที่เพื่อเพิ่มประสิทธิภาพในการจำแนกมาตรวัดน้ำที่ผิดปกติ โดยใช้ LSTM วิเคราะห์ปริมาณการใช้น้ำย้อนหลัง 49 เดือน และ XGBoost วิเคราะห์ข้อมูลมาตรวัดน้ำและปัจจัยเชิงพื้นที่ โดยทั้งสองแบบจำลองถูกนำมาฝึกเป็น Base Models และใช้ Stacking Ensemble รวมผลลัพธ์โดยใช้ Logistic Regression เป็น Meta-Model เพื่อทำนายว่ามาตรวัดน้ำผิดปกติหรือปกติ (Binary Classification) ซึ่งประสิทธิภาพของแบบจำลองถูกวัดโดย Confusion Matrix และ F1-Score

2. วิธีการดำเนินการวิจัย

2.1 พื้นที่ศึกษา

พื้นที่ศึกษาคือ การประปาส่วนภูมิภาคสาขารังสิต จังหวัดปทุมธานี ซึ่งมีจำนวนผู้ใช้บริการน้ำประปามากที่สุดของการประปาส่วนภูมิภาค โดยมีพื้นที่ให้บริการครอบคลุมพื้นที่ประมาณ 125,741 ตารางกิโลเมตร (อิงตามรัศมีจากโครงข่ายเส้นท่อประมาณ 100 เมตร) ให้บริการประชน 158,673 ครัวเรือน โดยมีปริมาณน้ำผลิตจ่าย 7,754,649 ลบ.ม./เดือน และมีปริมาณน้ำจำหน่าย 4,197,157 ลบ.ม./เดือน คิดเป็นปริมาณน้ำสูญเสียที่ 45.87% (ข้อมูล ณ กันยายน 2567)



รูปที่ 1 แผนที่พื้นที่ให้บริการน้ำประปาของ กปภ. สาขารังสิต

2.2 การรวบรวมและจัดเตรียมข้อมูล

2.2.1 ข้อมูลสำหรับฝึกและทดสอบแบบจำลอง

ชุดข้อมูลที่ใช้ในการศึกษาครั้งนี้ประกอบด้วยข้อมูลมาตรวัดน้ำที่มีการตรวจสอบสภาพการใช้งาน ณ สถานที่ติดตั้ง ในพื้นที่การให้บริการน้ำประปาของ กปภ.สาขารังสิต ตั้งแต่เดือน ตุลาคม พ.ศ. 2563 ถึง กันยายน พ.ศ.

2567 จำนวน 20,652 จุด และข้อมูลจุดตำแหน่งที่เคยมีการซ่อมท่อประปา จำนวน 48,099 จุด นำมาวิเคราะห์เพื่อสร้างข้อมูลความสัมพันธ์เชิงพื้นที่ระหว่างตำแหน่งของมาตรวัดน้ำและตำแหน่งที่มีการซ่อมท่อประปา ได้แก่ ระยะทางระหว่างตำแหน่งมาตรวัดน้ำกับตำแหน่งที่มีการซ่อมท่อประปาที่อยู่ใกล้ที่สุด (Nearest Distance) และ จำนวนจุดซ่อมท่อที่อยู่ใกล้เคียงมาตรวัดน้ำแต่ละเครื่องในรัศมี 100 เมตร (Buffer) โดยชุดข้อมูลมาตรวัดน้ำถูกแบ่งออกเป็นข้อมูลสำหรับฝึกและทดสอบแบบจำลองในอัตราส่วน 70:30 ซึ่งเป็นการใช้ข้อมูลส่วนใหญ่ในการฝึกแบบจำลองและแบ่งอีกส่วนไว้สำหรับประเมินประสิทธิภาพของแบบจำลอง

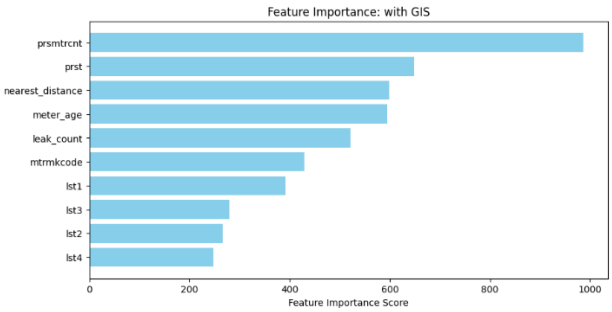
2.2.2 ปัจจัยที่ส่งผลต่อความเที่ยงของมาตรวัดน้ำ

การตรวจสอบและประเมินประสิทธิภาพของมาตรวัดน้ำ ณ สถานที่ติดตั้งใช้งาน ปัจจัยสำคัญที่ส่งผลต่อความคลาดเคลื่อนในการวัด ได้แก่ การชำรุดเสื่อมสภาพจากอายุการใช้งาน การหยุดชะงักของการไหล การเปลี่ยนแปลงลักษณะการไหล รวมถึงอุปกรณ์ใช้น้ำภายในครัวเรือนที่ถูกดัดแปลง ปัจจัยเหล่านี้ถือเป็นจุดที่อาจก่อให้เกิดความผิดพลาดในการวัดและยังมีปัจจัยภายนอก เช่น คุณภาพน้ำ อุณหภูมิ อากาศ ความผันแปรของแรงดันในระบบท่อ และอัตราการไหลที่เปลี่ยนแปลงตามพฤติกรรมการใช้น้ำของผู้บริโภค นอกจากนี้จากการศึกษายังพบว่าการซ่อมแซมท่อประปาอาจส่งผลให้เกิดการปนเปื้อนของเศษตะกอนไปสะสมในมาตรวัดน้ำ ซึ่งส่งผลต่อความเที่ยงของมาตรวัดน้ำ โดยปัจจัยที่ใช้ในการศึกษานี้ แสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลปัจจัยที่ใช้ในการฝึกแบบจำลอง LSTM และ XGboost

ข้อมูล	ประเภท	LSTM	XGboost
เส้นทางอ่านมาตรวัดน้ำ	ข้อมูลระเบียน		/
อายุการใช้งาน			/
ประเภทผู้ใช้น้ำ			/
ยี่ห้อมาตรวัดน้ำ			/
เลขอ่านหน่วยน้ำล่าสุด			/
หน่วยการใช้น้ำย้อนหลัง 48 เดือน (ก.ย.63-ต.ค.67)		/	/
หน่วยการใช้น้ำเดือนปัจจุบัน (ก.ย. 67)		/	/
สภาพมาตรวัดน้ำ		/	/
จำนวนจุดซ่อมท่อในรัศมี 100 เมตร	ข้อมูลเชิงพื้นที่		/
ระยะห่างจากจุดซ่อมท่อที่ใกล้ที่สุด			/

นอกจากนี้ยังมีการวิเคราะห์ค่าน้ำหนักของปัจจัยที่ใช้ในแบบจำลอง XGBoost โดยผลการวิเคราะห์ฟีเจอร์ 10 ลำแรกได้แก่ 1) เลขอ่านมาตรวัดน้ำล่าสุด (prsmtrcnt) 2) หน่วยการใช้น้ำเดือนล่าสุด (prst) 3) ระยะทางระหว่างตำแหน่งมาตรวัดน้ำไปยังตำแหน่งที่มีการซ่อมท่อประปา (nearest_distance) 4) อายุการใช้งานมาตรวัดน้ำ (meter_age) 5) จำนวนจุดซ่อมท่อในรัศมี 100 เมตรจากจุดติดตั้งมาตรวัดน้ำ 6) รหัสยี่ห้อมาตรวัดน้ำ (mtrmkcode) 7) หน่วยการใช้น้ำย้อนหลัง 1 เดือน 8) หน่วยการใช้น้ำย้อนหลัง 3 เดือน 9) หน่วยการใช้น้ำย้อนหลัง 2 เดือน และ 10) หน่วยการใช้น้ำย้อนหลัง 4 เดือน จะเห็นว่าปัจจัยเชิงพื้นที่ทั้งจำนวนจุดซ่อมท่อที่ใกล้เคียง และระยะห่างจากจุดซ่อมท่อที่ใกล้ที่สุด แบบจำลองให้ค่าน้ำหนักความสำคัญอยู่ในช่วงลำดับต้น ๆ แสดงดังรูปที่ 2



รูปที่ 2 ลำดับความสำคัญของข้อมูลที่ใช้ฝึกแบบจำลอง XGBoost

2.3 Ensemble machine learning algorithms

2.3.1 XGBoost (Extreme Gradient Boosting) method

XGBoost พัฒนาโดย Chen และ Guestrin (2016) ถูกออกแบบมาเพื่อใช้ในงานในหลายรูปแบบเช่น การจำแนกประเภท (Classification) การทำนายค่า (Regression) และ การจัดอันดับข้อมูล (Ranging) เป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งช่วยลดข้อผิดพลาดในการทำนาย มีอัลกอริธึมเป็น Ensemble Learning ซึ่งเป็นการนำตัวทำนายหลายตัว (multiple learners) มาทำงานร่วมกันเพื่อเพิ่มความแม่นยำของผลลัพธ์ โดยใช้เทคนิค Gradient Boosting เพื่อปรับแต่งการทำนายโดยการลดค่า loss function ผ่านวิธีการไล่ระดับ (Gradient Descent) นอกจากนี้ยังรวมเอาเทคนิคการ Regularization เพื่อลดความซับซ้อนของโมเดลและป้องกันการเกิด Overfitting เช่น การเลือกใช้เฉพาะบางฟีเจอร์ในแต่ละรอบของการเทรน (Column Subsampling) นิยมใช้ในงานที่เกี่ยวกับข้อมูลแบบโครงสร้าง เช่น การทำนายยอดขาย การตรวจจับการฉ้อโกง [9] ด้วยความสามารถดังกล่าวจึงนำ XGboost มาประยุกต์ใช้ในการจำแนกมาตรวัดน้ำที่มีแนวโน้มผิดปกติ ซึ่งมีสมการเป็นดังนี้

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (1)$$

โดย $l(\hat{y}_i, y_i)$ เป็น Loss Function ที่ใช้วัดข้อผิดพลาดระหว่างค่าแบบจำลองทำนายกับค่าจริง และ $\Omega(f_k)$ เป็น Regularization Term ที่ใช้ควบคุมความซับซ้อนของแบบจำลองเพื่อลดปัญหา Overfitting

2.2.3 Long short-term memory (LSTM) method

Long Short-Term Memory (LSTM) เป็นแบบจำลองการเรียนรู้แบบมีผู้สอน (Supervised Learning) ที่อยู่ในกลุ่มของโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Networks: RNNs) ซึ่งได้รับการพัฒนาขึ้นเพื่อแก้ไขข้อจำกัดของ RNN แบบดั้งเดิม โดยเฉพาะปัญหา vanishing gradient ซึ่งเกิดจากการที่ค่าเกรเดียนสำหรับการปรับค่าน้ำหนัก

(Gradient) ลดลงอย่างต่อเนื่องจนใกล้ศูนย์ในระหว่างกระบวนการถอยกลับเพื่อปรับค่าน้ำหนัก (Backpropagation) ส่งผลให้แบบจำลองไม่สามารถเรียนรู้ข้อมูลจากอดีตที่อยู่ไกลออกไปได้อย่างมีประสิทธิภาพ [10] LSTM มีโครงสร้างภายในที่ประกอบด้วย หน่วยความจำหลัก (Memory Cell) และกลไกควบคุมข้อมูล 3 ส่วนหลัก หรือที่เรียกว่า Gates ได้แก่ 1) Forget Gate ใช้ในการตัดสินใจว่าข้อมูลจากสถานะก่อนหน้าควรถูกลืมหรือไม่ 2) Input Gate ใช้ควบคุมการนำข้อมูลใหม่เข้าสู่สถานะปัจจุบัน และ 3) Output Gate ใช้ควบคุมข้อมูลที่จะถูกส่งออกจากหน่วยความจำไปยังขั้นตอนถัดไป ด้วยโครงสร้างที่สามารถรักษาข้อมูลสำคัญในลำดับเวลาได้อย่างมีประสิทธิภาพ LSTM จึงเหมาะสมอย่างยิ่งสำหรับการประมวลผลข้อมูลลำดับ (Sequential Data) เช่น การพยากรณ์ข้อมูลอนุกรมเวลา การรู้จำเสียงพูด หรือการวิเคราะห์พฤติกรรมการใช้น้ำย้อนหลังเพื่อคาดการณ์ความผิดปกติของมาตรวัดน้ำ ฟังก์ชันการทำงานของ LSTM สามารถนิยามได้ดังนี้

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

f_t คือค่าที่ออกจาก Forget Gate (มีค่าระหว่าง 0-1) เพื่อเลือกว่าจะเก็บค่าจากสถานะก่อนหน้าหรือไม่, σ เป็นฟังก์ชัน Sigmoid ที่ทำให้ค่าอยู่ระหว่าง 0 ถึง 1 , W_f คือเวคเตอร์น้ำหนักของ Forget Gate, h_t เป็น Hidden State ที่ใช้ส่งต่อข้อมูลไปยังขั้นถัดไป, h_{t-1} เป็นค่า Hidden State จาก timestep ก่อนหน้า, x_t คือข้อมูลอินพุตของ timestep ปัจจุบัน และ b_f คือค่าไบแอสของ Forget Gate

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \quad (3)$$

ซึ่ง C_t เป็นสถานะเซลล์ปัจจุบันที่ใช้เก็บข้อมูลระยะยาว (Cell state), $\odot$ เป็นการคูณแบบ element-wise C_{t-1} คือค่าของสถานะเซลล์ก่อนหน้า, i_t เป็นค่าที่ออกจาก Input Gate (มีค่าระหว่าง 0-1) เพื่อเลือกว่าจะเพิ่มค่าจาก Candidate State หรือไม่ และ $\widetilde{C}_t$ คือค่าของข้อมูลใหม่ที่อาจถูกเพิ่ม (Candidate Cell State)

2.2.4 Stacking ensemble (STK) method

Stacking Ensemble (STK) เป็นเทคนิคที่ใช้ในการรวมผลลัพธ์ของแบบจำลองที่เป็น Base Learners หลายตัวเข้าด้วยกัน แล้วนำไปผ่าน Meta-Learner เพื่อพยากรณ์ผลลัพธ์ขั้นสุดท้าย เพื่อช่วยเพิ่มความแม่นยำในการทำนายของแบบจำลอง [11] นอกจากนี้ยังพบว่าการใช้เทคนิค ML หลายตัวร่วมกันช่วยเพิ่มความแม่นยำในการตรวจจับความผิดปกติได้อย่างมีประสิทธิภาพ การใช้ STK เพื่อรวมผลลัพธ์จากแบบจำลอง Random Forest,Gradient-boosting, XGboost ให้ผลลัพธ์การจำแนกที่ดีกว่าการใช้แบบจำลองทีละตัว มักใช้สมการเชิงเส้นเป็นพื้นฐาน การศึกษานี้ใช้ Logistic Regression (LR) มีพื้นฐานการคำนวณดังสมการที่ (4) [12]

$$\hat{p} = \left(\frac{1}{1+e^{(\hat{\alpha}+\sum \hat{\beta}_i z_i)}}\right) \quad (4)$$

โดย $\hat{p}$ คือค่าความน่าจะเป็นของเหตุการณ์ ซึ่งหมายถึงโอกาสที่มาตรวัดน้ำจะมีความผิดปกติ, $\hat{\alpha}$ คือค่า Intercept ของแบบจำลอง, $\hat{\beta}_i$ คือค่าสัมประสิทธิ์ของตัวแปร z_i ซึ่งเป็นผลทำนายจาก XGBoost และ LSTM ที่ใช้เป็น Base Learners ในงานวิจัยนี้คือผล

2.4 การประเมินประสิทธิภาพแบบจำลอง

การศึกษานี้ใช้ข้อมูลสำหรับทดสอบคนละชุดกับข้อมูลตัวอย่างที่ใช้ในการฝึกแบบจำลอง เพื่อประเมินประสิทธิภาพแบบจำลองกับข้อมูลมาตรวัดน้ำที่มีการชำรุดจริงในช่วงเวลาภายหลังเดือนกันยายน 2567 โดยใช้ข้อมูลมาตรวัดน้ำที่ถูกสำรวจแล้วพบว่าเป็นมาตรวัดน้ำที่ผิดปกติ คือมีสภาพชำรุด มีความเที่ยงในการวัดปริมาตรน้ำมีความคลาดเคลื่อนเกินเกณฑ์ที่ยอมรับได้ของ กปภ. ซึ่งอิงตามมาตรฐานของสำนักงานชั่งตวงวัด (ชตว.) โดยเป็นการจัดทำตารางเปรียบเทียบระหว่างจำนวนที่ทำนายได้ถูกต้องและผิดพลาด ซึ่งเรียกตารางนี้ว่า “Confusion Matrix” แบ่งเป็น 4 ส่วนตามการจำแนกมาตรวัดน้ำที่มีความผิดปกติและปกติ ได้แก่ จำนวนที่ทำนายว่าผิดปกติและผิดปกติจริง (True Positive: TP), จำนวนที่ทำนายว่าปกติและปกติจริง (True Negative: TN) , จำนวนที่ทำนายว่าผิดปกติแต่จริง ๆ แล้วเป็นปกติ (False Positive: FP) และจำนวนที่ทำนายว่าปกติแต่จริง ๆ แล้วมีความผิดปกติ (False Negative: FN) ซึ่งใช้ในการคำนวณค่าต่าง ๆ สำหรับประเมินประสิทธิภาพของแบบจำลองได้เป็นอย่างดี [13] ได้แก่ ความแม่นยำ (Precision), ความสามารถในการตรวจจับความผิดปกติ (Recall), อัตราส่วนของการจำแนกมาตรวัดน้ำผิดปกติและปกติได้อย่างถูกต้องเมื่อเทียบกับจำนวนทั้งหมด (Overall Accuracy: OA) มีค่าระหว่าง 0-1 โดยหากมีความถูกต้องสูงค่าจะเข้าใกล้ 1 และ F1-score ซึ่งเป็นตัวชี้วัดสำคัญในการประเมินประสิทธิภาพของแบบจำลอง ML เนื่องจากเป็นค่าที่รวม Precision และ Recall เข้าด้วยกัน ซึ่งมีการคำนวณค่าต่าง ๆ ตามสมการที่ (5)-(8)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (8)$$

3. ผลการศึกษาและอภิปรายผล

การศึกษานี้ ได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองที่ใช้ในการจำแนกมาตรวัดน้ำที่มีความผิดปกติ ในขั้นตอนการพัฒนาแบบจำลองได้ทดสอบค่า Hyperparameter ของแบบจำลอง LSTM และ XGboost โดยแบบจำลอง LSTM ได้รับการปรับแต่งค่าพารามิเตอร์ (Hyperparameter tuning) ด้วยเทคนิค Bayesian Optimization เพื่อให้ได้โมเดลที่มี

ประสิทธิภาพดีที่สุด และใช้ Grid SearchCV ในการค้นหาค่าพารามิเตอร์ที่ดีที่สุดของแบบจำลอง XGBoost แสดงรายการในตารางที่ 2

ตารางที่ 2 ชุดค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับการฝึกแบบจำลอง ML

แบบจำลอง	พารามิเตอร์	คำอธิบาย	ค่า
LSTM	num_layers	จำนวนชั้น LSTM	1
	units_0	จำนวนหน่วย (neurons) ใน LSTM ชั้นแรก	160
	activation	ฟังก์ชันกระตุ้นที่ใช้ในแต่ละหน่วย	tanh
	recurrent_dropout	อัตราการ dropout สำหรับการเชื่อมต่อ recurrent	0.1
	dropout_0	อัตราการ dropout ของ LSTM ชั้นแรก	0.4
	optimizer	อัลกอริทึมที่ใช้ปรับค่า weight	adam
	learning_rate	อัตราการเรียนรู้	0.005
XGboost	booster	ประเภทของตัวเรียนรู้	gbtree
	colsample_bytree	สัดส่วนของฟีเจอร์ที่สุ่มเลือกมาใช้ในแต่ละต้นไม้	0.8
	learning_rate	อัตราการเรียนรู้	0.05
	max_dept	ความลึกสูงสุดของแต่ละต้นไม้	7
	min_child_weight	น้ำหนักรวมขั้นต่ำของลูก node ถ้า node ใดค่าน้อยกว่านี้จะไม่แบ่งต่อ	5
	n_estimators	จำนวนต้นไม้ทั้งหมดที่สร้าง (หรือรอบที่ boost)	400
	subsample	สัดส่วนของข้อมูล train ที่สุ่มใช้ในแต่ละต้นไม้	0.8
Stacking	final estimator	แบบจำลองสุดท้าย (meta-model)	logistic regression
	stack_method	วิธีการที่ base models ใช้ในการสร้าง feature สำหรับ final_estimator	predict_proba

ในการประเมินประสิทธิภาพของแบบจำลองใช้ตัวชี้วัด 4 อย่างได้แก่ Precision, Recall, F1-score และ OA ของแบบจำลองแต่ละแบบจำลองในการจำแนกกลุ่มมาตรวัดน้ำที่ผิดปกติได้อย่างถูกต้อง โดยผลลัพธ์แสดงในตารางที่ 3 เมื่อเพิ่มฟีเจอร์เชิงพื้นที่ (Spatial Features) เข้าไปในแบบจำลอง XGBoost พบว่าประสิทธิภาพเพิ่มขึ้นอย่างมีนัยสำคัญ โดยค่า F1-score เพิ่มจาก 83.37% เป็น 85.40% และ OA เพิ่มเป็น 85.25% แสดงให้เห็นว่าข้อมูลเชิงพื้นที่มีประโยชน์ในการช่วยให้แบบจำลองสามารถจำแนกกลุ่มมาตรวัดน้ำที่ผิดปกติได้แม่นยำขึ้น ซึ่งแสดงให้เห็นว่า ความหนาแน่นของของจุดซ่อมท่อและความใกล้เคียงกับจุดซ่อมท่อ มีความสัมพันธ์กับความผิดปกติของมาตรวัดน้ำ

ตารางที่ 3 ประสิทธิภาพของแบบจำลองในการจำแนกข้อมูลชุดทดสอบ

แบบจำลอง	Precision	Recall	F1-score	OA
LSTM	88.67%	88.66%	88.61%	88.61%
XGboost	83.37%	83.13%	83.16%	83.24%
XGboost with GIS	85.40%	85.17%	85.21%	85.25%
Stacking (LSTM+ XGboost)	88.31%	88.30%	88.31%	88.32%
Stacking (LSTM+ XGboost with GIS)	90.03%	90.01%	90.02%	90.03%

การรวมแบบจำลอง LSTM และ XGBoost (Stacking) โดยไม่ใช้ข้อมูลเชิงพื้นที่ได้ค่า OA 88.32% และ F1-score 88.31% ซึ่งดีกว่าแบบจำลองแบบเดี่ยวเล็กน้อย แต่เมื่อเพิ่มข้อมูลเชิงพื้นที่เข้าไปในแบบจำลอง XGBoost และใช้เทคนิคการรวมแบบจำลอง (Model Stacking) ร่วมกับ LSTM แบบจำลองนี้ให้ผลลัพธ์ดีที่สุดทุกตัวชี้วัด โดยมีค่า OA เท่ากับ 90.03% และ F1-score เท่ากับ 90.02% ซึ่งสอดคล้องกับงานวิจัยของ Kanyama et al. (2024) ที่ระบุว่าการผสานเทคนิคการเรียนรู้ของเครื่องหลายแบบร่วมกันสามารถเพิ่มความแม่นยำของแบบจำลองได้ [12] และสนับสนุนผลการศึกษาของ Intarat et al. (2024) ที่ใช้การรวมแบบจำลองด้วยวิธี Stacking ระหว่าง XGBoost, Gradient Boosting และ Random Forest ซึ่งช่วยเพิ่มประสิทธิภาพของแบบจำลองได้อย่างมีนัยสำคัญ [13]

4. สรุปและข้อเสนอแนะ

ผลการศึกษานี้แสดงให้เห็นว่าปัจจัยต่าง ๆ เช่น อายุการใช้งาน เลขหน่วยการใช้น้ำสะสม ประวัติการใช้น้ำ ระยะห่างจากจุดซ่อมท่อ และความถี่ในการซ่อมท่อที่อยู่ใกล้กับจุดติดตั้งมาตรวัดน้ำ มีความสำคัญในการคาดการณ์แนวโน้มที่มาตรวัดน้ำจะมีสภาพการทำงานผิดปกติซึ่งส่งผลต่อความเที่ยงในการการวัดปริมาตร เมื่อเทียบกับแบบจำลองอื่น ๆ ที่ใช้ในการศึกษานี้ แบบจำลอง Stacking ที่ใช้ข้อมูลเชิงพื้นที่ในการฝึกแบบจำลองให้ประสิทธิภาพสูงสุดในการจำแนกมาตรวัดน้ำที่มีความผิดปกติ ซึ่งมีค่าความถูกต้องรวมสูงสุดที่ 90.03% อย่างไรก็ตามยังมีมาตรวัดน้ำที่ยังไม่ได้รับการตรวจสอบในภาคสนาม ณ สถานที่ติดตั้ง จำนวนมาก การทดสอบความเที่ยงต่อจุดติดตั้งใช้ระยะเวลานาน พื้นที่ศึกษามีขนาดใหญ่ จึงถูกจำกัดด้วยเวลาและงบประมาณ จึงควรมีศึกษาเพิ่มเติมเพื่อหาปัจจัยอื่น ๆ ที่อาจส่งผลต่อความเที่ยงของมาตรวัดน้ำ เพื่อเพิ่มความแม่นยำในการจำแนกมาตรวัดน้ำที่มีความผิดปกติด้วยการเรียนรู้ของเครื่องต่อไป

กิตติกรรมประกาศ

ข้าพเจ้าขอแสดงความขอบคุณอย่างสูงไปยังผศ.ดร.ชัยยุทธ์ เจริญผล อาจารย์ที่ปรึกษา และคณาจารย์ภาควิชาวิศวกรรมสำรวจ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำและความคิดเห็น

ต่อเนื่องตลอดระยะเวลาการทำวิจัยนี้ซึ่งช่วยให้การวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณทีมงานกองมาตรวัดน้ำ กองเทคโนโลยีระบบจำหน่าย สำนักเทคโนโลยีน้ำสูญเสีย งานระบบภูมิสารสนเทศ กองเทคโนโลยีระบบประปาและการประปาส่วนภูมิภาคสาขารังสิต ที่ให้ความช่วยเหลือในด้านการเก็บข้อมูลและการวิเคราะห์ข้อมูลต่าง ๆ ที่มีประโยชน์ต่อการศึกษาครั้งนี้

ขอขอบพระคุณ การประปาส่วนภูมิภาค ที่ให้การสนับสนุนทุนการศึกษา และทรัพยากรที่จำเป็นในการทำวิจัยนี้

ขอขอบพระคุณอย่างยิ่งต่อครอบครัว บุคคลใกล้ชิด ญาติสนิท และมิตรสหายทุกท่าน ที่ได้ให้กำลังใจและสนับสนุนผู้วิจัยด้วยความเมตตาและห่วงใยตลอดระยะเวลาของการศึกษาค้นคว้าและดำเนินการวิจัยในครั้งนี้

เอกสารอ้างอิง

[1] USAID, *The Manager's Non-Revenue Water Handbook for Africa*, The United States Agency for International Development, 2010.

[2] การประปาส่วนภูมิภาค,เอกสารรายงานจำนวนมาตรวัดน้ำประจำเดือน กันยายน ปีงบประมาณ 2567, 2567.

[3] การประปาส่วนภูมิภาค, มาตรฐานการปฏิบัติงานกระบวนการบริหารจัดการน้ำสูญเสีย (เชิงพาณิชย์), 2567.

[4] มรุพัชร จำนงค์วงศ์, "การลดน้ำสูญเสียในระบบประปา," The Journal of KMUTNB, vol. 26, no. 9, 2559.

[5] S. Nofal, A. Alfarrarjeh, and A. Abu Jabal, "A use case of anomaly detection for identifying unusual water consumption in Jordan," Water Supply, vol. 22, no. 1, pp. 1131–1140, 2021.

[6] J. C. Carrasco-Jiménez, F. Baldaro, and F. Cucchietti, "Detection of anomalous patterns in water consumption: An overview of approaches," in Intelligent Systems and Applications, Cham: Springer, 2021, pp. 19–33.

[7] M. Carratù, S. D. Iacono, G. D. Leo, V. Gallo, C. Liguori, and A. Pietrosanto, "Smart water meter based on deep neural network and undersampling for PWNC detection," IEEE Trans. Instrum. Meas., vol. 72, pp. 1–11, 2023.

[8] M. N. Kanyama, F. Bhunu Shava, A. M. Gamundani, and A. Hartmann, "Machine learning applications for anomaly detection in smart water metering networks: A systematic review," Phys. Chem. Earth A/B/C, vol. 134, p. 103558, 2024.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., San Francisco, CA, USA, 2016, pp. 785–794.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[11] M. N. Kanyama, F. Bhunu Shava, A. M. Gamundani, and A. Hartmann, "Machine learning applications for anomaly detection in smart water metering networks: A systematic review," Phys. Chem. Earth A/B/C, vol. 134, p. 103558, 2024.

[12] K. Intarat, P. Yoomee, A. Hussadin, and W. Lamprom, "Assessment of landslide susceptibility in the intermontane basin area of Northern Thailand," Environ. Nat. Resour. J., vol. 22, no. 2, pp. 158–170, 2024.

[13] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, New York, NY, USA: Springer, 2013, ch. Evaluating Predicted Classes, p. 269. doi: 10.1007/978-1-4614-6849-3.